RANDOM MOTION FOR CAMERA CALIBRATION

Zoltán Szlávik[2]*, Tamás Szirányi*[1,2]*,László Havasi* [1]*, Csaba Benedek*[1,2]
[1]Péter Pázmány Catholic University,
H- 1083 Budapest Práter u. 50/a., Hungary,
e-mail: {havasi,}@digitus.itk.ppke.hu
[2]Analogic and Neural Computing Laboratory, Hungarian Academy of Sciences,
H-1111 Budapest, Kende u. 13-17, Hungary,
e-mail: {szlavik, bcsaba, sziranyi}@sztaki.hu

ABSTRACT

In the paper we show that by using co-motion statistics random motion can be used for the registration of views and calibration of cameras. The introduced algorithm finds point correspondences in two views without searching for any structures and without the need for tracking continuous motion.

1. INTRODUCTION

Calibration of cameras and registration of different views is a basic task for many applications, such as: stereovision, 3-dimensional reconstruction, tracking across multiple views [6] [8][9].
Usually an algorithm for the alignment of different views and calibration of cameras has the following steps:

1. Feature detection;
2. Extraction of candidate point-pairs;
3. Rejection of outliers and estimation of the model that does the alignment;
4. Alignment of different views;
5. Estimation of epipolar geometry.

In the paper we will present an algorithm for the first four steps of the above general schema. Having the point correspondences extracted the estimation of epipolar geometry and calibration of cameras can be done by well-known algorithms [1][8][10].
Matching different images of a single scene may be difficult, because of occlusion, aspect changes and lighting changes that occur in different views. Still-image matching algorithms [2][3][4][5] search for still features in images such as: edges, corners, contours, color, shape etc. They are usable for image pairs with small differences; however they may fail at occlusion boundaries and within featureless regions. They may fail if the chosen primitives or features cannot be reliably detected. The views of the scene from the various cameras may be very different, so we cannot base the decision solely on the color or shape of objects in the scene.
In a multi-camera observation system the video sequences recorded by cameras can be used for estimating matching correspondences between differentviews. Video sequences in fact also contain information about the scene dynamics besides the static frame data. Scene dynamics is an inherent property of the scene independently of the camera positions, the different zoom-lens settings and lighting conditions. References [6] and [7] present approaches in which motion-tracks of the observed objects are aligned. However, in these cases a robust capability for object tracking is assumed; and this is the weak point of both methods.
As a previous work the use of co-motion statistics for the estimation of projective geometry was introduced in [9][10]. The approach proposed in [10] is an extension, albeit a considerable one, of the previously mentioned sequence-based image matching methods for non-structured estimation [6][7]. In [9][10] we have introduced the use of co-motion statistics for the matching and alignment of two overlapping views and estimation of the common groundplane. In that approach, instead of the trajectories of moving objects, the statistics of concurrent motions – the so-called co-motion statistics – were used to locate matching points in pairs of images. The inputs of the system are video sequences derived from cameras located in fixed positions; however, the actual camera positions, orientations, and zoom settings are unknown.
The main advantage of the use of co-motion statistics that no *a priori* information about motion, objects or structures is needed. The disadvantage of co-motion statistics is that the system needs huge memory for storing it. In [11] we presented an algorithm for the efficient estimation of co-motion point-pairs and a robust feature extraction method. In which less memory is needed for coding scene dynamics, the calculations have been done on-line.
Here we show that not only deterministic motions [11], but also random motions can be used for the extraction of point correspondences and estimation of epipolar geometry. The paper is organized as follows: in section 2 main steps of the algorithm are described, subsection 2.1 is a brief summary about co-motion statistics, which is followed by the description of change detection and coding of scene dynamics, extraction of point-correspondences; in section 3 the experiments and results are presented.

2. ESTIMATION OF EPIPOLAR GEOMETRY

The algorithm described here is based on the use of co-motion statistics for matching images [10]. The steps of the algorithm are the following:

1. Detect changes.
2. Store changes and the dynamics of the scene that the scene can be reconstructed later.
3. Extract point-correspondences from the stored scene dynamics – detection of features, extraction of candidates.

4. Estimate the model – rejection of outliers and estimation of fundamental matrix.

2.1. Co-motion statistics

Scene dynamics is encoded in co-motion statistics, so if static features (corners, edges etc.) cannot be reliably detected the information for matching can be extracted from co-motion statistics [9][10].
In case of single video sequence a motion statistical map for a given pixel can be recorded as follows: when motion is detected in a pixel, the coordinates are recorded of all pixels where motion is also detected at that moment. In the motion statistical map the values of the pixels at the recorded coordinates are updated. After all, this statistical map is normalized to have global maximum equal to 1.
In case of stereo video sequences to each point in the images, two motion-statistic maps are assigned: a local and a remote. Local map means the motion-statistical map in the image from the pixel is selected, the remote motion-statistical map is refer to the motions in the other image. After the motion detected on the local side, for the points defined by the local motion map the local statistical map updated by the local motion map. For each point where motion is detected on the local side, the local motion map of the remote side updates the corresponding remote statistical map. An example of co-motion statistics for inlier point-pairs can be seen in Figure 1.

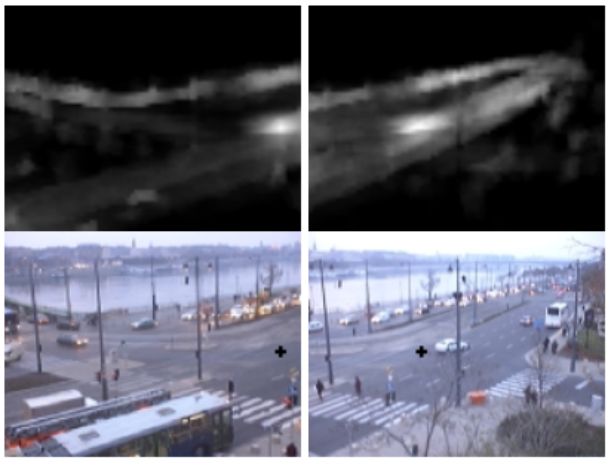

Figure 1 Top images: example of co-motion statistics for inlier point-pairs. Below: a corresponding point-pair is shown in the images of the left and right cameras.

The advantage of this interpretation of the scene dynamics is that point correspondences in the above case were interpreted as maximums of statistical maps and their extraction is very simple. The main disadvantage of co-motion statistics is that the system must keep two statistical maps (grayscale pictures) for each pixel of input image, which means that the algorithm needs huge memory, in case of 160*120 statistical map resolution it means 1,4 GBs!

2.2. Change detection and coding scene dynamics

For the detection of changes we have used the absolute difference of two consecutive frames. This method is fast and very sensible with low threshold value. The result of the change detection can be either a binary (pixel value is 1 if change is detected and 0 else) or a grayscale (pixel value is the real value of absolute difference) image. Some disadvantage comes from the cases that often detect noises and background flashings. In our algorithm for matching of images we do not need precise change detection and object extraction, because of the later statistical processing these minor errors are irrelevant.
Having the result of change detection the scene dynamics can be coded and stored. To overcome the problem that huge memory is needed for storing co-motion statistics we propose to store the motion history in a vector for each pixel instead of storing an image-size map for each of them as in [9][10]. This motion history vector has as many entries as long is the video sequence and in each of its entry has 1 if change was detected at the given frame or zero if not. This coding reduces the memory needs while the scene dynamics is also coded in the vectors. The disadvantage is that for the extraction of point correspondences all the motion histories of cameras must be compared. The advantage of this method is that less memory is needed for the storage of scene dynamics than in the case of co-motion statistics: 2.3 MB in case of 500 frame long image sequence, while for the storing of co-motion statistics 1,4 GB is needed.

2.3. Extraction of point correspondences

Usually the estimation of point correspondences in two given images consists of three steps. Firstly, features are detected then candidates of point pairs are extracted and, finally the outliers are rejected and the given model is estimated.

2.3.1. Feature detection
From the images of the two views we extract feature points related to pixels of real objects (cars, people etc.) moved through them. We don't want to extract pixels in which change was detected due to flashings or random noise on the background. For the extraction of these points we have compared two methods. In the first method we are integrating the motion histories. If this value is above some threshold then the corresponding pixel is selected as a feature point. This method is very sensitive to the threshold value.
In the second we have calculated the Shannon entropy of pixels' motion history vectors.

$$entropy = -\sum p(x_i)\log p(x_i) \qquad (1)$$

where $p(x_i)$ is the frequency of x_i in vector v, v – real-valued motion history vector. Experiments with different indoor and outdoor videos showed that the entropy of motion history vectors of flashings and other random noise and the entropy of motion history of deterministic motion of real objects have Gaussian distribution as it is shown in Figure 2.

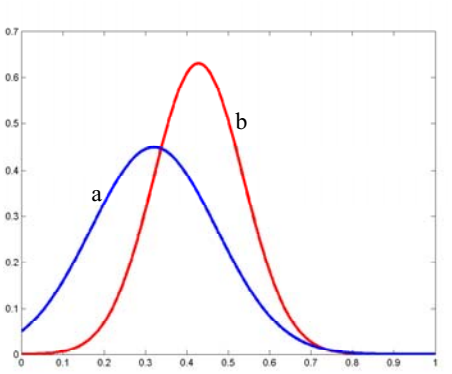

Figure 2 The distribution of Shannon entropy of flashings (a curve) and deterministic motion (b curve).

In order to get the right threshold value for the entropy we analyzed the obtained distributions. The proportion of outliers in the set of point candidates is essential for the RANSAC algorithm that we have used for the rejection of outliers (see section 2.4.3) and estimation of fundamental matrix [8]. Larger the proportion of outliers larger is the running time of the RANSAC algorithm [8]. The aim of the feature extraction is to provide a necessary amount of matchings for the estimation of fundamental matrix (at least seven matchings). Setting the threshold value equal to 0.2 usually provides enough matchings and it is easy to calculate that the proportion of outliers in the set of point-candidates will be 15%, which is small enough to ensure small running time of the RANSAC algorithm [8].
Instead of traditional definition of entropy for vector v, we have also tested the formula for the estimation of the "entropy":

$$entropy^* = -\frac{1}{\log(N)}\sum v_i \log v_i \qquad (2)$$

where v_i are the elements of the history vector v, N – the length of history vector v.

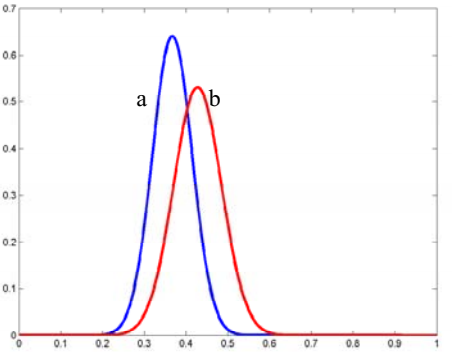

Figure 3 The distribution of *entropy** (see formula (2)) of flashings (a curve) and deterministic motion (b curve).

Applying this formula is not a serious restriction to our algorithm. The meaning of its output is similar to that of the output of the traditional formula. If in a given pixel the system is observing only small flashings then the value of $entropy^*$ will be high (logarithm of a small number is a large number in absolute value). If object is moving through the pixel then the detected change will be much higher then in case of flashings and the value of $entropy^*$ will be low (logarithm of a large number is small). Figure 3 shows the distribution of *entropy** of flashings and deterministic motions. Similarly, we set the threshold value for *entropy** equal to 0.32 as for Shannon entropy. The main advantage of formula (2) against (1) that it can be calculated on-line, from frame to frame as the implemented change detection algorithm.
It is obvious that if all our candidate points are from the same region of input images and close to each other then small error in point coordinates (which comes from the change detection, which is, of course, not perfect) will result in great error in final alignment of the whole images. To reduce it we forced points to be better distributed in the region by introducing some structural constraints: images are divided into blocks of n*n and for each block the algorithm searches for only one candidate point, for which the integrate of motion history is the maximum and its entropy is within a given interval.

2.4.2 Extraction of candidate point pairs

Having the features points detected in both views for the extraction of candidate point-pairs the feature points of different views must be compared. For the comparison of feature points, the corresponding motion history vectors in our case, we have implemented different methods for binary and real-valued motion history vectors.
In the case of real-valued motion history vectors the extraction of candidate point pairs is based on the calculation of the correlation between a given feature point and feature points of the other view.
In the case of binary motion history the time-series of the history-vectors are filtered. This morphological filter removes single peaks and groups neighbor peaks if they are within a predefined distance. After filtering the Hamming distance is calculated as correlation between two binary motion history vectors of different views.

2.4.3. Robust estimation of the model and rejection of outliers

The geometry of two views is well understood [8]. In the case of uncalibrated cameras the fundamental matrix encodes the relationship between two views [8][10]. For the estimation of fundamental matrix and rejection of outliers from the set of candidate point-pairs we have implemented the RANSAC algorithm [8][10]. Having the matchings extracted and the fundamental matrix computed the camera matrices easily can be estimated and 3D locations can be reconstructed by using triangulation [8].

3. EXPERIMENTAL RESULTS

In order to show that random motion can be used for camera calibration we set up the following experiment. A small tree was blown with a periodically rotating (forth and back) ventilator and the generated random motion of the tree's leaves was recorded with two cameras at resolution 320×240, at same zoom level and with same cameras (LAB videos). The point correspondences were extracted by using real-valued motion history and thresholding of integrated motion history for feature

extraction. The above described entropy-based feature extraction cannot be used for feature extraction in this experiment because of the random motion of tree's leaves. The estimated epipolar pencil can be seen in Figure 4.

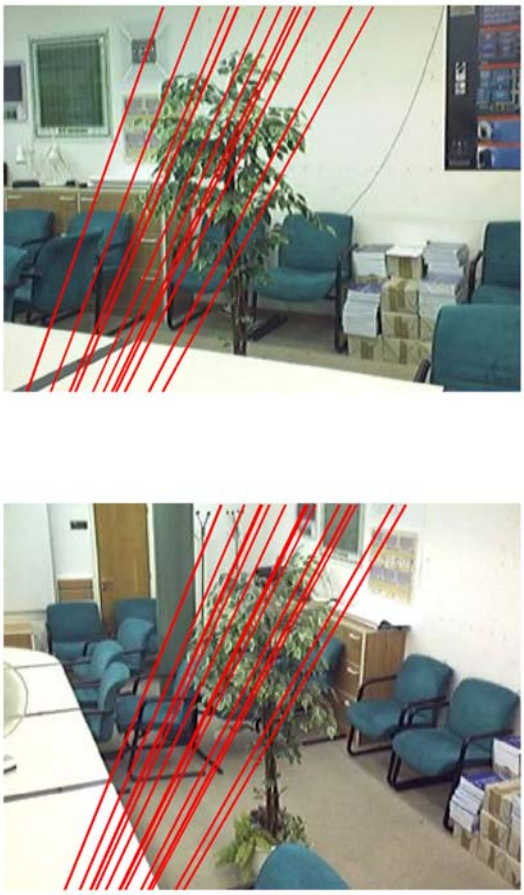

Figure 4 The epipolar pencils for the LAB test videos.

4. CONCLUSIONS

We have shown that partially overlapping camera views can be registered by motion history vectors of images' reference pixels of outdoor cameras placed in freely-chosen positions, viewing arbitrary scenes where motion is present, and this matching is automatic without any human interaction. We have shown that the registration of views can be done even if cameras record random motion. Based on the registration of co-motion point-pairs we have estimated the epipolar geometry of the scene and cameras easily can be calibrated by using well-known methods [8].

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge the support received from the NoE MUSCLE project of the EU.

6. REFERENCES

[1] O. D. Faugeras, Q.-T. Luong, S. J. Maybank, "Camera self-calibration: Theory and experiments," in *Proc. ECCV '92, Lecture Notes in Computer Science*, vol. 588, Berlin Heidelberg New York, Springer-Verlag, pp. 321-334, 1992.

[2] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence Journal*, vol. 78, pp. 87-119, 1995.

[3] S. T. Barnard, W. B. Thompson, "Disparity analysis of images," *IEEE Trans. PAMI*, vol. 2, pp. 333-340, 1980.

[4] J. K. Cheng, T. S. Huang, "Image registration by matching relational structures," *Pattern Recog.*, vol.17, pp.149-159, 1984.

[5] J. Weng, N. Ahuja, T. S. Huang, "Matching two perspective views," *IEEE Trans. PAMI*, vol. 14, pp. 806-825, 1992.

[6] L. Lee, R. Romano, G. Stein, "Monitoring activities from multiple video streams: establishing a common coordinate frame," *IEEE Trans. PAMI*, vol. 22, 2000.

[7] Y. Caspi, D. Simakov, and M. Irani, "Feature-based sequence-to-sequence matching," in *Proc. VAMODS (Vision and Modelling of Dynamic Scenes) workshop, with ECCV'02*, Copenhagen, 2002.

[8] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge,Cambridge University Press, 2003.

[9] Z. Szlávik, L. Havasi, T. Szirányi, "Estimation of common groundplane based on co-motion statistics", in *Proc. of ICIAR'04*, *Lect. Notes in Computer Science,* pp. 347-353, 2004.

[10] Z. Szlávik, L. Havasi, T. Szirányi, "Image matching based on co-motion statistics", Proc. of 2nd Int. Symposium on 3DPVT, Thessaloniki, 2004.

[11] Z. Szlávik, T. Szirányi, L. Havasi, Cs. Benedek, "Optimizing of searching co-motion point-pairs for statistical camera calibration", ICIP'05.